DOCUMENT RESUME

ED 155 733 CS 501 974

AUTHOR Ellis, Donald G.
TITLE Issues in Analyzing Sequential Interaction Data: A Plea for Rigor in Matters of Observation.
PUB DATE Dec 77
NOTE 33p.; Paper presented at the Annual Meeting of the Speech Communication Association (63rd, Washington, D.C., December 1-4, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Communication (Thought Transfer); Group Behavior; Human Relations; *Interaction Process Analysis; *Interpersonal Relationship; *Research Design; *Research Methodology; Social Relations; *Standards
IDENTIFIERS *Communication Research

ABSTRACT

A great need exists for increased attention to standards of rigor in collecting sequential interaction data. As in any other research strategy, powerful statistics are not sufficient to organize data meaningfully. A knowledge of individuals, their relational histories, and communication are all necessary to insure proper conclusions about the conceptual nature of the interaction. Coding systems must preserve function; statistical structure is less important than the qualitative nature of a coding scheme. Research must observe the following standards of quality. Before a coding scheme is used it should reflect considerable agreement about the presence and definability of a category. Coders should not be aware of the research purposes. Demographics and compositional variables are important, as is the setting where people interact. Researchers should be wary of comparisons made across units of analysis, and must give special attention to within-group versus across-group comparisons. There should always be a match between concept and method. (References are included.) (DF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED155733

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ISSUES IN ANALYZING SEQUENTIAL INTERACTION DATA: A PLEA FOR RIGOR IN MATTERS OF OBSERVATION

Donald G. Ellis
Department of Communication
Purdue University
West Lafayette, IN 47907

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Donald G. Ellis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Presented to Speech Communication Association
Washington, D.C., December, 1977

CS501 974

ISSUES IN ANALYZING SEQUENTIAL INTERACTION DATA:
A PLEA FOR RIGOR IN MATTERS OF OBSERVATION

I sense a renewed interest in sequential interaction data and analysis. In the 1940's when Shannon and Weaver (1948) and Miller and Frick (1949) formulated information theory for studying response patterns, researchers were excited about the potential applications to human interaction. In the years to come Cherry (1957), Chomsky (1957), Attneave (1959), Khinchin (1957), and Quastler (1958) would refine and generalize these issues to various academic disciplines. However, after the late 1950's these procedures were eclipsed by the growing popularity of classical parametric statistics.

The renewed interest in sequential analysis is stimulated by a series of conditions. First, many people in communication, whether they express it or not, have the feeling that we are going nowhere. They see volumes of research but little knowledge. Eventhough the producers and consumers of research reports remain active and enjoy their work, they cannot shake the feeling that these bits and pieces of "knowledge" do not fit. And telling yourself that someday a general theory will emerge, which will make sense of everything, is of little consolation. This feeling of aimlessness is partially a function of our inability to link actual communicative behaviors with the constructs or theoretical terms we have formulated. That is, we have ignored what should be our most precious scholarly possession--talk. It is humorous, bordering on sad, how little attention we devote to this most defining characteristic. The correspondence rules which link theoretical terms with observations have been concerned with self-reports in the form of questionnaire observations only; however the observational phenomena which give our theoretical terms cognitive significance should be observable

verbal interaction. Although some scholars would express the matter differently, it is essentially this failure of adequate correspondence between theory and observation which accounts for the feeling that much of our research has little application to actual social interaction.

Moreover, there is renewed research activity on social interaction in related disciplines. As early as 1953 Cartwright (p. 422) observed that it was "really surprising how much of social psychology is in the form of verbal behavior." Although verbal behavior as we define it is not relevant to other fields, most social sciences encounter some form of social interaction. Developmental psychologists are paying more attention to infant-caretaker interaction (Lewis and Rosenblum, 1974). Altmann (1965) has examined social communication in primates, and Wilson (1976) studies signal systems across species. Raush et. al. (1974) study family interaction and clinical psychologists in general display increased concern with social interaction.

In short, more people realize that interaction is the stuff of life and not paper-and-pencil reconstructions. Still, researchers who attempt some form of systematic observation may find themselves confronted with a bewildering array of decisions about strategies and data which their experience and training has not equipped them to deal with. The purpose of this paper is to expose the reader to issues in observing and analyzing sequential interaction. What follows is a general discussion of systematic observation, coding systems, and data analysis. However, I will pay particular attention to certain issues in coding data.

Moreover, alternatives for data analysis will be an important part of this paper. Since many of these issues are discussed in detail elsewhere, my goal is to provide insights into basic issues, point up new procedures, and direct the reader to other sources.

Systematic Observation

The decision to use systematic observation should, of course, be based on the theoretical concerns of the research. Weick (1968), Riley (1963), Baumrind (1964), and Holsti (1969) offer occassions and reasons for using observational methods. First, systematic observation provides a certain type of data, data which captures the temporal connection of "whole" events. Many other research procedures result in a snapshot of personal attributions made by a subject. These data are often time bound and incapable of yielding moment-to-moment sequences. Second, systematic observation is appropriate when subject self reports are either distorted or incapable of identifying the phenomenon of interest. Some phenomena are subtle and escape the attention of the subject; others are completely alien to the subjects cognitive structure. At this juncture some may criticize the creation of data which has no place in the phenomenological field of the subject. Some researchers have been criticized for creating esoteric and unique catergories of interaction. Suffice it to say that utility and basic issues of realiability and validity will determine the ultimate value of catergories of interaction. Moreover, the social sciences are replete with constructs used to explain human behavior which have no meaning for the layman. Ask, for example, one of your students whether he has an internal or an external locus of control.

The conceptualizations implied by a series of questions designed to measure "locus of control" are no different when applied to catergories of interaction. Finally, observation methods should be employed when the researcher wants to correlate meaning with behavior. Theoretical constructs explain or interpret verbal interaction; however many of these constructs are only loosely associated with actual verbal behavior. For example, systematic observation methods would prove essential for determining the communicative correlates of locus of control.

The relationship between the observer and the observational system is a central aspect of collecting sequential interaction data. The human observer armed with an observational system can be the most useful and important element in gathering data or a debilitating source of error. Weick (1968), Brandt (1972), Dean, Eichhorn, and Dean (1967), and Barker and Wright (1955) discuss the influence of the trained observer. These authors detail the importance of observer influence, procedures for the observer to employ when collecting data, how to minimize observer effect, and how to measure, manage, and control the observers inferences.

The point I wish to emphasize is that the observer is engaging in <u>controlled perception</u>. The definition of systematic observation includes the selection and recording of events which are consistent with empirical goals. The researcher is interested in specific behavioral events and sets about the task of determining how these events will be defined and observed. The adequacy and conceptual utility of definitions is more important than the potential for observer error.

If theoretical constructs are adequately defined, then training an observer is nothing more than practicing controlling his perception. The possibility for unwarranted observer inference is lessened when definitions and conceptualizations are rigorous and precise.

Bochner (1977) criticizes the tendency to digitally classify verbal events and maintains that it is easy to "train naive judges to see the same acts similarly but uncorrectly." Yet training naive observers to see acts similarly is exactly the goal of the researcher. This is the nature of reliability and ensures accurate selection and interpretation of phenomena. The referent for "accurate" is the researcher's conceptual and operational definitions. If they are pooor the observer's results will be poor. If these definitions perserve the functions of the interaction and observers are well trained, then the data will be valuable.

CODING SYSTEMS

Most of what I have said underscores the importance of the coding schemes used in systematic observation. The researcher's data are only as good as his category system and no amount of sophisticated data analysis will improve the conceptual quality of the results. Researchers should pay particular attention to category development, units of analysis, reliability, and validity.

Development

Berelson (1952), Borgatta (1962), Medley and Mitzel (1963), Heynes and Zander (1966), Weick (1968), Holsti (1969), and Brandt (1972) provide cogent discussions of category development. These authors begin by emphsizing the importance of specificity. A category must be rigorously defined and apply to a single and clearly identifiable phenomenon. The molar-molecular nature of the phenomenon becomes important since increases in molarity are

associated with a loss of specificity. Duncan and Fiske (1977) label categories recognition rules or rules for the identification of events. The fully explicit recognition rule enables the observer to first note the occurrence of the event. Once this is accomplished the observer can count, time, or measure the event in some desired manner.

Few communication scholars--except some in the area of nonverbal communication--enjoy fully explicit recognition rules. It is simple, for example, to explicitly code a head nod, the blink of an eye, or a turn at talk: however more molar or meaning-laden categories such as "shows agression," "seems friendly," or "gives information" are rife with potential for error. The category can become increasingly meaningless with a diversity of codifiable units. Nevertheless, until categories with molar characteristics are of no interest to communication researchers, we should continue to pursue their explicit recognition rules. A diversity of behaviors which can be classified into a single category is important and desirable if the link between the specific behaviors and the category is sound. A theory, for example, which generates and confirms predictions about new substantive areas is more general and possesses a higher degree of confirmation. A molar category of interaction can be considered analogous. As the number of legitimate behavioral instances which can be coded into the category increases, the category becomes more generally applicable with a higher degree of confirmation. Again, the crucial issue is not the molarity of the category but the correspondences between the category and behavioral instances.

The problem with most observational analysis in communication research is that few, if any, researchers devote enough time to their category system. Hawe's (1975) analysis of medical interviews

is one exception because he developed the categories from the intended population of interaction. However, I could find no instances of where communication researchers doing interaction analysis provided explicit empirical referents for their categories of interaction. A typical example is provided by Wood (1977) who was studying adaptive behavior by leaders. One of her categories is labeled task and is defined tautologically as "issues involved in discussion." Examples are gaining information or opinions, clarifying ideas, or evaluating data. These provide a nice constituitive definition but do not render the category significant. Nowhere in the definition is the category linked to empirical events. Moreover, terms like information, opinion, and data require the same explicit definitions. This is an instance where Bochner's (1977) warning about naive judges coding reliably but incorrectly becomes a reality. Although the coders in this research were highly reliable (.93) they leave a host of issues about the categories unsolved. For example, if the author has solved the issue of identifying information in interaction streams, then I wish she would share it with us. I know of no one who "gets at" the matter of pragmatic information. Moreover, what is the difference between information and data? When does clarifying an idea cease and result in a new idea, or information, or data, or whatever these distinctions are meant to be? Future efforts to code verbal behavior must proceed with considerably more rigor.

Finally, category systems have traditionally been developed as mutually exclusive and exhaustive systems. Very simply, an event was classified into one and only one category and the set of categories presumbly classified all behaviors (see Brandt, 1972).

The rule of exclusivity has weakened lately, largely because of the belief that a behavioral instance is multidimensional and can serve a variety of functions. While the assumption of multifunctions is sensible, it is also very dangerous. To argue that a verbal act contains multiple meaning requires the researcher to specify each meaning context. This problem is not insurmountable but will certainly tax the researchers ability to draw distinctions. Secondly, there is the assumption that each behavior in a category is equivalent to the others in the category. If, however, one event possesses multiple meaning and is coded into a second category, the data analysis is problematic. It simply makes no sense to statistically consider the one event commensurate with the others. However, the assumption of equivalence within a category poses problems because one loses the intensity and significance of a selected event. Wright (1960) maintained that the relationship between the significance of an event and rate of occurrence may be inverse. A well placed and timely "display of emotion," or "act of aggression" may be rare but important. This significance will be lost if the event ends up as merely a tally on a score sheet. In short, nonexlusive coding schemes might be desirable but play havoc with the complexities involved in data analysis and interpretation. Mann (1967) and Purcell and Brady (1965) offer some examples of how behavioral events can be assigned to several categories.

Category systems should be exhaustive. Exhaustiveness implies that all related behaviors within the research context can be classified into one of the categories. While the requirement for exclusivity can be relaxed, exhaustiveness is desirable.

When the researcher is developing a coding scheme for a particular research interest he must account for all "related" behaviors. This could become a conceptual problem, however the committed researcher must do all he can to encompass the behavorial system which characterizes the phenomena of interest. Exhaustiveness should be considered a requirement for systemic complexity. That is, even if the researcher is interested in a single specific behavorial event a complete characterization or explanation of the event will include related topics. So the researcher working on gaze behaviors, for example, will not only consider gaze direction, time, and rate but will ultimately be concerned with a series of concurrent behaviors e.g. gaze while speaking, eye movement, back channeling, gestures etc. To consider a coding scheme exhausitive is essentially to determine system boundries. When units of information are overlooked or classified into residual categories the data are very difficult, if not impossible, to interpret.

Unit of Analysis

The unit of analysis defined by a coding scheme is, again, a function of the research interest. Furthermore, the issues involved in the various units of analysis are related to the above section on the correspondence between categories and data. Categories which are essentially low level constructs with direct ties to data (e.g., "who follows whom") often dictate the nature of the analytical unit, although constructs at any level of generality influence the selection of an analytical unit.

Bakeman and Dabbs (1976) and Bakeman (1977) discuss approaches to specifying units of analysis. Assuming that observers reliably record codes which make conceptual sense to the researcher, behavior patterns can be examined as sequential or concurrent. Moment-to-moment

changes become sequential patterns; multiple behaviors occuring at a given moment form concurrent patterns. For example, an interactant could greet, smile, and extend his hand in a sequential pattern, or do all three simultaneously in a concurrent pattern. A second distinction concerns event or time bases. An event is a functional behavioral unit such as smiling, agreeing, or speaking; a time base is a durational measure such as 15 seconds of greeting, 2 seconds of smile etc. The combination of these distinctions produces four data types.

Event sequential. In this case the observer defines functional behavioral units and notes sequences or pattern development. This is where categories are usually mutually exclusive and defined so only one event may occur at a time. Most interaction analysis in communication is event sequential.. Ellis (1976), for example, found that decision-making groups "initiate information," "extend information," reinforce, "and initiate information." This type of data is prevalent in classroom interaction research (Flanders, 1966), mother-child interaction (Stern, 1974), and animal behavior (Altmann, 1965).

Event concurrent. Again, the researcher ignores duration and codes predefined function units but these units can occur concurrently. Many behaviors occur simultaneously but these are often difficult to code and pose even more statistical problems. Exactly what constituties an adequate behavior complex is difficult to formulate. Most investigators measure duration and transform the data into one of the types mentioned below.

Time sequential. Here one measures duration but events may occur only one at a time. An adequate theoretical reason for choosing one time metric over another is often a problem. Why the observer should code every second, or two seconds, or three minutes should be more than a matter of convention. Hawes (1972) codes medical interviews every ten seconds but provides no reason for choosing this particular metric. To code according to a time interval necessarily implies that the time interval is important. One loses the contiguous nature of functional events since either more than one event occurs in the time interval or the interval is too short for the event to occur. Sequential analysis makes little sense if the researcher is losing information about the phenomenon because it is lost in a time interval.

Time concurrent. This is where duration is measured but events can overlap. Time concurrent data is useful and important because of its flexibility. A member of a dyad, for example, might take 10 seconds to utter an agreement statement and at the same time the observer could note various body or facial movements. An example appears in Brown et. al. (1975).

These analytical units have statistical implications which will be discussed below. I would add that these data types are flexible and certainly capable of transformation from one type to another. Though members of a dyad elicit different behaviors concurrently, it is possible to code the same behaviors sequentially paying attention to one person, then the other, and so on. Time series analyses are more available when data are displayed sequentially. A more pressing issue, however, is the nature of a coding schemes reliability.

Reliability and Validity

The critical issue in reliability is agreement among observers. A coding scheme that can be learned and applied with consistency is reliable. Guetzkow (1950) argues for unitizing reliability which focuses on the correct number of classifiable units and categorizing reliability which pretains to correct assignment to categories. There are four sources of poor reliability. The first three are research design such as inadequate data sampling; poor training for observers; and the coding scheme itself (e.g., too few categories, poor category definition). The fourth problem area is statistical tests for reliability.

Reliability is the extent of agreement beyond chance expectations. The notion of "beyond chance expectations" is important because many reliability coefficients fail in this respect. Simple percentage of agreement, for example, is a very poor index of reliability because chance expectations are not accounted for. Another dimension of reliability, which is related to chance expectations but often overlooked by researchers, is the importance of the number of categories. By coding two categories we would expect 50% agreement by chance alone. Using three, four, or five categories obviously changes the expected levels of agreement. Therefore an adequate measure of reliability must account for chance, consider the number of categories in the coding system, and avoid computational practices which create sensitivities to particular data types.

Incidentally, Cohen (1960) makes a case for the triviality of testing reliability coefficients for statistical significance. Calculating a standard error and referring the obtained statistic to a normal curve yields little information. To conclude that the result is significantly different from zero is not very enlightening because we should expect so much more of our reliability.

The largest dragon I would like to slay is the use of Pearson r as a test for reliability. The procedure has serious problems, some of which can be very misleading, and should be avoided. First, and least important, Pearson r requires at least interval level data (Siegel, 1956, p. 195). Since coding units of behavior into categories yields nominal level data, this requirement is violated.

Secondly, Pearson r is very sensitive to categories with large frequencies. When a few categories have a high proportion of acts assigned to them these categories have a disproportionate influence on the r statistic. A ten category system which contains two frequently and easily codible categories will inflate the r statistic and give a false sense of reliability. While coders may be reliable with respect to the frequent categories, they could be unreliable on the remaining eight categories. The r statistic will drop drastically if computed on the eight categories only. Pearson r is capable of lulling the researcher into a sense of reliability when, in fact, he is unreliable on 80% of the categories.

The third limitation is crucial. The purpose of the reliability statistic is to indicate the level of agreement between two coders. The Pearson r does not measure agreement but assesses linear relationships. Robinson (1957) puts the matter succinctly when he writes that two variables which agree are correlated, but variables can be correlated and not necessarily agree. The data in Table I make the point. The r between the two sets of values is 1.0; the reliability is poor. One observer's scores are a linear function of the others. This is interesting but not an indication of coder reliability.

TABLE I

	Observer	
Category	A	B
1	50	96
2	37	83
3	36	82
4	18	41
5	35	58

The same is true for Spearman's Rho or Kendall's Tau. These statistics are versions of product-moment corelations and possess the same shortcomings as Pearson r. Rho is not a nonparametric index. Although tau is nonparametric, it is still a measure of the similarity between rank orders and not an index of agreement.

More acceptable measures of reliability are Scotts Pi (1955), Cohen's (1960) kappa, and Guetzkow's (1950) categorizing reliability. Scott's pi is a simple index of reliability which corrects for the number of categories and frequency with which they are used. Since an expected agreement figure is computed for each category, the statistic is not overly sensitive to large frequencies in a few categories. Cohen's kappa statistic is also directly interpretable as the proportion of joint agreement after chance agreement is removed. It is possible to compute a standard error for kappa and refer the result to the normal curve.

I would argue for the superiority of Guetzkow's estimate of reliability. Guetzkow begins with procedures for estimating unitizing reliability. Coders may divide the units at different points or disagree on how many units are in a segment. Guetzkow's formula provides the researcher with an estimate of how often he

will error according to the masses of data he codes. Under the assumption that classifying a unit is independent from the previous or subsequent classification, Guetzkow's formula indicates the affect on theoretical agreement as K (N of categories) increases. Agreement among coders is determined by the correct selection of a category and obtained agreement is an estimate of theoretical agreement. One can calculate the lower and upper limits for various magnitudes of t and determine the level of confidence in the statistic. Guetzkow (1950) contains all of the necessary computations and Hawes (1972) offers an example.[1]

Validity is equally as important for systematic observation as any other research strategy. The issues are the same to the extent that the researcher should be concerned about content, predictive, concurrent, and construct validity. However, too many people using interaction analysis techniques rely on face or content validity. They decide that their categories measure what they want them to measure and then stop. Determining predictive or concurrent validity for particular coding systems is not very difficult. If, for example, predictions are made about how individuals should interact then applying a coding scheme to determine these patterns will help clarify the validity of the coding scheme. Ellis (1978) coded interaction and then used the results to place subjects into a group formed on the basis of relational interaction (see Ellis, 1976; Farace & Rogers, 1975). A discriminant analysis yielded trait profiles which were associated with a particular interaction strategy. In this way the coding schemes and concept of relational interaction gained some predictive and concurrent validity.

DATA ANALYSIS

It would be impossible to survey very many types of data analysis. I would like to comment on a few tradition al procedures and suggest some that many people may be unfamiliar with. For some reason, I suspect that people are skeptical about systematic observation because of untraditional statistical methods. Observation data are typically nominal or composed of frequency data and not very amendable to parametric statistics without some minimal data modification. Statistical tests and assumptions appropriate for nominal data cast into one of the four data types above, are usually unfamiliar. The first test discussed is the most common--chi square.

Assuming a matrix of frequency data which represents content or contiguous states, the chi square statistic is an appropriate procedure. Nevertheless, the statistic detects significant differences very easily (especially as N increases), yields relatively little information, and can be replaced by more useful statistics. The largest drawback to the chi square is its limited information yield. After one finds significant constraint in a matrix the chi square stops. Magnitudes of effects, reductions in uncertainty, descriptive qualities of sequences, and comparisons between matrices which represent time periods or the result of an experimental manipulation, are all difficult to uncover.

Many people using chi square statistics on contingency tables are unaware of procedures for partitioning contingency tables to find specific sources of variance. A significant chi square for any contingency table beyond a 2 X 2 is simply an indication that "somewhere" observed proportions do not match theoretical proportions. Yet I could not find a single instance in our journals where contingency tables were partitioned correctly. In brief, it is

possible to partition a chi square into additive components analogous to orthogonal contrasts. This partitioning results in subdividing a contingency table into (R-1) (C-1) 2 X 2 tables, each with one degree of freedom. A 2 X 3 contingency table, then, would have (R-1) (C-1) or (2-1) (3-1) or two contrasts. Chi squares performed on these contrasts would sum to the overall chi square and indicate specific effects. The reader should see Chochran (1954), Castellan (1965), Armitage (1955), Irwin (1949), and Kimball (1954) for more about partitioning contingency tables.

The most interesting aspect of the analytic procedures I describe below is their infrequent use by communication scholars. None of what follows is new; yet few people are familiar with them. Furthermore, many of the procedures involve computations which are not difficult to learn. I will mention various ways to approach data analytic problems with sequential data and provide original or helpful citations whenever possible.

Transition probabilities. Assume two people are talking and we are interested in coding their interaction from a relational perspective (see Ellis, 1976, 1978; Farace & Rogers 1975; Parks, 1977). The ↑ code is an attempt to dominate the relational definition; the ↓ is a submissive position; and → is the result of relational equality. Assume this dyad produced 311 relational messages and the frequencies are as follows;

$$F \quad [129 \qquad 146 \qquad 36]$$

Within the raw frequencies $f_2 = 146$ and is the number of times the ↓ (submissiveness) was encoded. (See Hewes, 1975). However, the lag1 transition frequency matrix is more interesting (See Bakeman & Dabbs, 1976; Sackett, 1974, 1977).

$$F\ (T) \qquad \begin{matrix} & & (T+1) & \\ \end{matrix}$$

$$\begin{bmatrix} \uparrow & 27 & 46 & 31 \\ \downarrow & 119 & 17 & 22 \\ \rightarrow & 10 & 22 & 16 \end{bmatrix}$$

Now f_{23}= 22 and is the frequency with which → follows ↓ . Observe that 311 observations produce 310 transitions because the first and last units have no antecedent or subsequent state respectively.

The above matrix can be converted into a lag 1 transition matrix with each cell representing the probability of one state following another.

$$P = (T) \qquad \begin{matrix} & & (T+1) & \\ \end{matrix}$$

$$\begin{bmatrix} .26 & .44 & .30 \\ .75 & .11 & .14 \\ .21 & .46 & .33 \end{bmatrix}$$

The probability of any ij transition is simply the cell frequency divided by the total row frequency.

These probability matrices can also be displayed graphically. (see Ellis, 1977; Hawes & Foley, 1973 Bakeman & Brown, 1977). Figure I graphically represents the above transition matrix

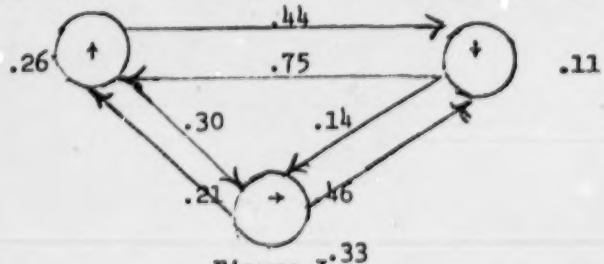

Figure I

The above emphasizes the descriptive characteristics of the data. It is important to ask, however, whether or not a transition is occurring more often then we would expect by chance. For example, if these states are independent then the probability of → following → (→→) should be no different than the probability of an → anywhere. The difference between observed and predicted values can be determined with the following binomial test. (See Bakeman and Dabbs, 1976; Scakett, 1977).

$$20 \quad z = (X-NP)/\sqrt{NPQ}$$

The X = the joint frequency of the cell in question. For the →→ cell X = 16. NP is the predicted joint frequency (N = frequency of → ; P =probability of →). NP = 36 X .115 = 4.16. NPQ is the variance of the difference between predicted and observed (36 X .115 X .89 = 3.52). Therefore:

$$Z = (16-4.167)/\sqrt{36 \times .115 \times .885}$$

$$= 11.83/1.914 = 6.18$$

These data are problematic with respect to statistical significance. There is considerable argument about the conceptual independence of behavioral events. A case can be made for either side. However, I maintain that we should consider behavioral events (in the form of communicative events) dependent upon one another. To say that when two people talk one person's statement is independent from the previous statement should be a definitional impossibility. "Theoretically" one can respond to another in any manner he desires. However, this does not adequately represent the actual communication around us which is the product of purposeful conjoint activity that becomes patterned over time. Initial research might examine the extent to which communicative events depart from pure chance. But as data bases develop from ongoing research, theoretical probabilities should rest on past research results and not chance expectations. For these reasons Bakeman and Dabbs (1976) suggest the Z score be considered an index and not to assign a p value to it.

Transition matrices, while important to Markov models, are not analogous to Markov models. These matrices can be used to describe data or perform any number of statistical operations. Markov models, however, are based on a set of assumptions which are either appropriate for characterizing data or not. Transition matrices must fit the assumptions of Markov models.

Profiles. Gottman and Bakeman (1977) also discuss how lag conditional probabilities can be profiled. For event sequential data, positions on the abscissa represent events following some criterion behavior, and the ordinate indicates probability. For example, the interaction produced by a dyad is profiled in figure 2. The solid line represents the probability of each event for the second person with person one encoding a ↑ as the criterion behavior. The dotted line is the same except the second person encoding a one up is the criterion behavior and the event probabilities are for the first person. The graph suggests the nature of the relationship between the two people. When the first person attempts to control the definition of the relationship, he usually meets with success as the second individual is most likely to comply (↓). However note what follows when person two attempts a ↑ ; he is very likely to be challenged by a ↑ from the first interactant (.75) thereby competing for control.

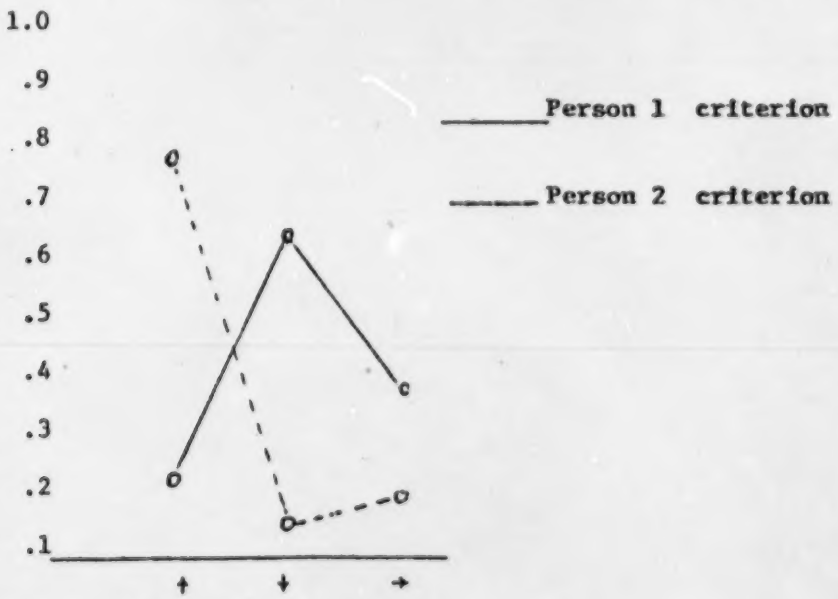

Figure 2

Detecting common sequences is also informative. Sackett (1977) and Gottman and Bakeman (1977) present the researcher with two alternatives. He can compute the probabilities of all possible chains (see Bobbitt, Gourevitch, Miller & Jensen, 1969), or examine lag sequential probabilities. Computing all possible chains for the three relational categories would result in $3^2 + 3^3$ or 36 possible chains. (Imagine the possibilities with 8 or 9 categories). A better approach involves designating each behavior as the criterion and tabling the probability of other behaviors at successive lags. If an event is sequentially independent of the criterion, then its conditional probability should be about the same as its simple probability. In this example, any deviation from .33 is a departure from expectation and this departure can be indexed with the Z score described above.

		lag		
	Behavior	1	2	3
A Criterion = ↑	↑	.30	.65*	.20
	↓	.60*	.15	.65*
	→	.10	.20	.15

		lag		
B Criterion = ↓	Behavior	1	2	3
	↑	.55*	.15	.75*
	↓	.30	.70*	.05
	→	.15	.15	.20

		lag		
C Criterion = →	Behavior	1	2	3
	↑	.30	.10	.25
	↓	.30	.30	.20
	→	.40	.60	.55

Table II

In Table 2 after a ↑ the most likely relational definition at lag 1 is a ↓ , a return to ↑ at lag 2 and back to ↓ at lag 3. Examining panel B supports our hunch that the data represent sequences of complementary relational definitions. When ↓ is the criterion behavior, a ↑ is most probable at lag 1, another ↓ at lag 2, and back to ↑ at lag 3.

Identifying structure. The most common measures of structure are information statistics. Since these procedures are common I will not elaborate. (See Attneave, 1959; Fisher, Glover, and Ellis, 1977; Hawes & Foley, 1976). Yet, there is one measure which is less familiar and seems to have potential.. This measure is the index of stereotypy. Applications of the index of stereotypy appear in Altmann (1965), Ellis (1976), Fisher, Glover, Ellis (1977), and Lenk (1977).

Stereotypy (S) details the structure at successive levels of order in sequential data. Specific sequences of data can range from completely determined by knowledge of preceeding events to completely independent of preceeding events. The index of stereotypy tells us where between these two extremes our data falls. This is a process of comparing each conditional uncertainty with its maximum value (See Altmann, 1965, p 512-513).

$$S(y) = 1 - \frac{U(y)}{\max U(y)}$$

$$Sx(y) = 1 - \frac{Ux(y)}{\max Ux(y)}$$

$$Sw,x(y) = 1 - \frac{Uw,x(y)}{\max Uw,x(y)}$$

and so on for higher orders.

Imagine a fourth order approximation for the relational interaction mentioned above. The stereotypy is illustrated in Figure 3.

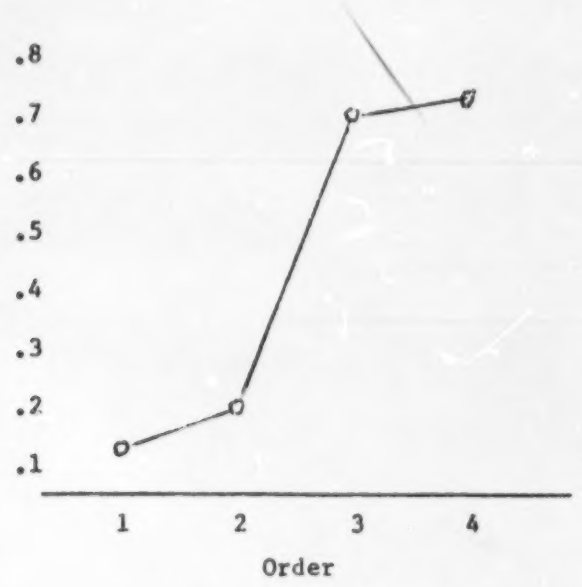

Figure 3

Knowing a previous event does little to reduce our uncertainty because the difference between S_1 and S_2 is slight. However, at third order approximations S increases sharply and indicates that knowing two preceeding events is quite informative and reduces uncertainty considerably. The increase at level four is small so we would conclude that patterning in these data occurs at three-event sequences.

Another example will demonstrate how stereotypy can be used to test for differences between experimental conditions. Stereotypy cannot be referred to a sampling distribution and assigned a p value; however the index can be very useful when data are sequential and dependent. Assume that subjects were separated into two groups based on cognitive complexity. One group is composed of highly complex subjects and the other group contains subjects scoring low on cognitive complexity. Each group must generate persuasive strategies one of which will be successful. Both groups begin

with a strategy which is not necessarily predictable so S_1 is barely above zero (see Figure 4). But both groups choose a second strategy which is similiar to the first. The first choice therefore affected the second and produces an increase in stereotypy (a substantial S_1-S_2 difference). However complex subjects, unlike noncomplex subjects, were less likely to choose a third strategy similar to the first, and so S_3 is higher for complex subjects. Indeed complex subjects requre half as many strategies before finding a successful one. This is a case where stereotypy indicates successive reductions in uncertainty as subjects make sequentially related choices.

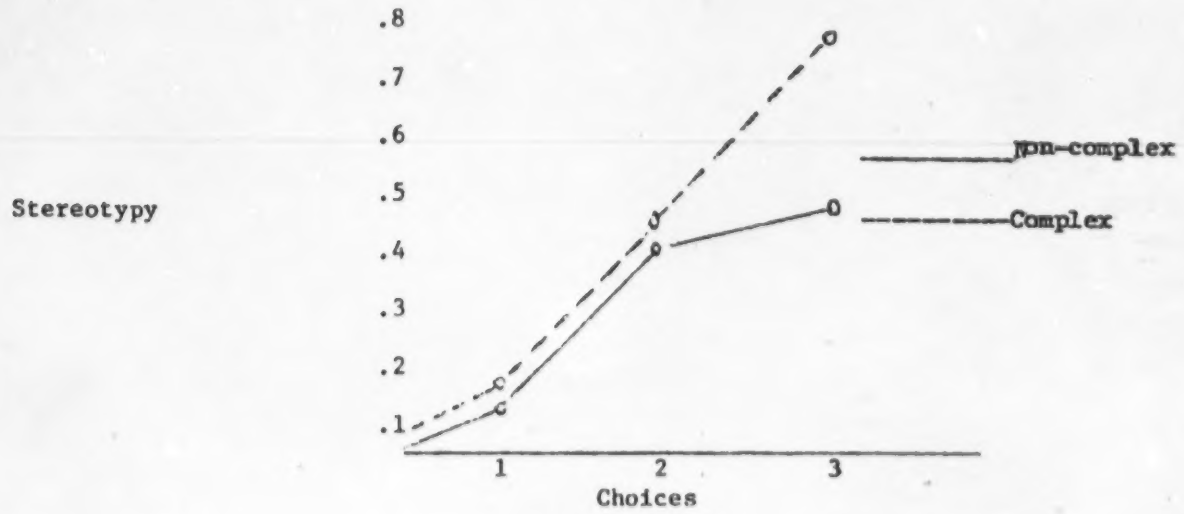

Figure 4

CONCLUSIONS

In short, this paper argues for increased attention and standards of rigor for sequential interaction data. The issues in sequential interaction data collection and analysis rest heavily on the conceptualizations of the researcher. Like any other research strategy, powerful statistics are not enough to organize data meaningfully. A knowledge of individuals, their relational histories, and communication are all necessary to ensure intelligent conclusions. An understanding of predictable streams of behavior is an exciting thought.

But knowing that one stream of interaction is more predictable than another does little to inform us about the conceptual nature of this interaction. Rosen (1972) maintains that coding systems must be function-perserving. That is, statistical structure is less important than the qualitative nature of a coding scheme.

Future researchers must demand increased rigor. As of now there are too many coding schemes purporting to measure the same thing, too few replications, too many poorly analyzed sets of data, and too little detailed discussion of how to construct and apply coding schemes. Research might bear in mind the following standards of quality.

1. Before a coding scheme is used it should reflect considerable agreement about the presence and definability of a category. Replications and multiple methods of measurement are crucial.

2. Coders should not be aware of the research purposes. Haley (1972) illustrates the extent of experimenter bias when group differences are known. Interaction differences between normal and schizophrenic families, for example, are less obvious when coders know nothing about the families.

3. Demographics and compositional variables are important. Jacob (1974), and Sanva (1961) found differences between families which differed on sociocultural factors.

4. McGrath and Altman (1966), and Mischel (1968) argue for the importance of accounting for the setting where people interact. Mischel found artificial differences between groups which were a function of different interaction settings.

5. Beware of comparing across units of analysis. There is a difference between the social and the psychological contexts of interaction. Some researchers impute motivations to individuals on

the basis of descriptive interaction while other know too little about the social interaction context. Some combination of procedures will prove most useful.

6. Be careful of within vs. across group comparisons. The processes within a group will contaminate comparisons between groups. Talk time, the ratio of one statement to another, or speaker order will influence between group comparisons.

7. Finally, there should always be a match between concept and method. Sometimes examining the simple frequency of occurrence of an act is insufficient. On the other hand, transition probabilities are not always necessary. Researchers should consider statistical control whenever possible. Percentages or analysis of covariance procedures can correct for potentially contaminating variables. (See Hadley, 1973).

I have been concerned with issues in sequential interaction data. I believe this reflects an important dimension of the current state of communication theory. Earlier I argued that many communication scholars were experiencing a sense of helplessness. I would add that there is an emerging breed of researchers who are disenchanted with traditional researcher practices. They realize that social interaction involves patterns which recur and combine in a variety of ways. Hopefully, some of the issues addressed above will prompt thought and action. To this extent, communication theory must advance.

REFERENCES

Altmann, S.A. Sociobiology of rhesus monkeys. II. Stochastics of social communication. Journal of Theoretical Biology, 1965, 8, 490-522.

Armitage, P. Tests for linear trends in proportions and frequencies. Biometrics, 1955, 11,375-386.

Attneave, F. Applications of information theory to psychology. New York: Henry Holt, 1959.

Bakeman, R., & Dabbs, Jr. Social interaction observed: Some approaches to the analysis of behavior streams. Personality and Social Psychology Bulletin, 1976, 2, 335-345.

Bakeman, R. Untangling streams of behavior: Sequential analyses of observation data. In G.P. Sackett (Ed.), Observing behavior, Volume 2: Data collection and analysis methods. Baltimore, University Park Press, in press.

Bakeman, R., & Brown, J.V. Behavioral dialogues: An approach to the assessment of mother-infant interaction. Child Development, 1977, 48, 195-203.

Barker, R.G., & Wright, H.F. Midwest and its children: The Psychological ecology of an American town. Evanston, Ill. Row, Peterson, 1955.

Baumrind, D. Naturalistic assessment of parent-child interaction. Paper presented at Conference on Research Methodology of Parent-Child Interaction, Syracuse, N.Y. 1964. Cited in Weick.

Berelson, D. Content analysis in communication research. Glencoe, Illinois: Free Press, 1952.

Bobbitt, R.A., Gourevitch, V.P., Miller, L.E., & Jensen, G.D. Dynamics of social interactive behavior: A computerized procedure for analyzing trends, patterns, and sequences. Psychological Bulletin, 1969, 71, 110-121.

Bochner, A.P. On taking ourselves seriously: An analysis of some current trends in interpersonal communication inquiry. Paper presented at the International Communication Association, Berlin, Germany, 1977.

Borgatta, E.F. A systematic study of interaction process scores, peer and self-assessments, personality and other variables. Genetic Psychology Monographs, 1962, 65, 219-291.

Brandt, R.M. Studying behavior in natural settings. New York; Holt, Rinehart, and Winston, 1972.

Brown, J.V., Bakeman, R., Snyder, P.A., Fredrickson, W.T., Morgan, S.T., & Hepler, R. Interactions of black inner-city mothers with their newborn infants. Child Development, 1975, 46, 677-686.

Cartwright, D.P. Analysis of qualitative material. In L. Festinger and D. Katz (eds.), Research Methods in the Behavioral Sciences. New York: Holt, 421-470.

Castellan, N.J. On the partitioning of contingency tables. Psychological Bulletin, 1965, 64, 330-338.

Cherry, C. On Human Communication. Cambridge, Mass.: M.I.T. Press, 1957.

Chomsky, N. Syntactic Structures. The Hague: Mouton, 1957.

Cochran, W. G. Some methods of strengthening the common χ^2 tests. Biometrics, 1954, 10, 417-451.

Cohen, J.A. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Dean, J.P., Eichhorn, R.L., & Dean, L.R. Observation and interviewing. In J.T. Doby (Ed.), An introduction to social research New York: Appleton-Century-Crofts, 1967.

Duncan, S., & Fiske, D. Face-to-face interaction. Lawrence Erlbaum, 1977.

Ellis, D.G. Relational interaction in decision-making groups. Paper presented to the ICA, Portland, Oregon, 1976.

Ellis, D.G. Trait predictors of relational control. In Brent Rueben (Ed.), Communication Yearbook II, in press.

Ellis, D.G. A social system model of relational control in two ongoing group systems. Paper presented to SCA, Washington, D.C. 1977.

Farace, V.V., & Rogers, E.L. Analysis of relational communication in dyads: New measurement procedures. Human Communication Research, 1975, 1, 222-239.

Flanders, N.A. Interaction analysis in the classroom: A manual for observers. Ann Arbor: School of Education, University of Michigan, 1966.

Gottman, J., & Bakeman, R. The sequential analysis of observation data. Unpublished manuscript, 1977.

Guetzkow, H. Unitizing and categorizing problems in coding qualitative data. Journal of Clinical Psychology, 1950, 6, 47-50.

Hadley, T., & Jacob, T. Relationship among measures of family power, Journal of Personality and Social Psychology, 1973, 27, 6-12.

Haley, J. Critical overview of present status of Family interaction research. In J. Framo (Ed.), Family Interaction: A Dialogue Between Family Researchers and Family Therapists. New York: Springer, 1972.

Hawes, L.C., & Foley, J. Group decisioning: Testing a finite stochastic model. In G. Miller (Ed.), Explorations in Interpersonal Communication, Beverly Hills, Calif.: Sage, 1976.

Hawes, L.C. & Foley, J. A Markov analysis of interview communication. Speech Monographs, 1973, 40, 208-219.

Hawes, L.C. Development and application of an interview coding system. Central States Speech Journal, 1972, 23, 92-99.

Hawes, L.C. The effects of interviewer style on patterns of dyadic communication. Speech Monographs, 1972, 39, 114-123.

Hewes, D.E. Finite stochastic modeling of communication processes: An introduction and some basic readings. Human Communication Research, 1975, 1, 271-283.

Heynes, R.W., & Zander, A.F. Observation of group behavior. In L. Festinger & D. Katz (Eds.), Research methods in the behavioral sciences. New York: Holt, 1966.

Holsti, O.R. Content analysis for the social sciences and humanities. Reading, Mass.: Addison-Wesley, 1969.

Irwin, J.O. A note on the subdivision of X^2 into components. Biometrika, 1949, 36, 130-134.

Jacob, T. Patterns of Family conflict and dominance as a function of child age and social class. Developmental Psychology, 1974, 10, 1-12.

Khinchin, A.I. Mathematical Foundations of Information theory. New York: Dover, 1957.

Kimball, A.W. Short-cut formulas for the exact partition of chi-square in contingency tables. Biometrics, 1954, 10, 452-458.

Lenk, D. A stochastic examination of communication development in self-analytic groups. Doctoral dissertation. Dept. of communication, Temple University, 1977.

Lewis, M., & Rosenblum, L.A. (Eds.), The effect of the infant on its care giver. New York: John Wiley, 1974.

Mann, R.D. Interpersonal styles and group development. New York: Wiley, 1967.

McGrath, J.E. & Altman, I. Small group research. New York: Holt, 1966.

Medley, D.M. & Mitzel H.E. Measuring classroom behavior by systematic observation. In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Miller, G.A., & Frick, F.C. Statistical behavioristics and sequences of responses. Psychological Review, 1949, 56, 311-324.

Mischel, W. Personality and Assessment. New York: Wiley, 1968.

Parks, M.R. Relational communication: Theory and research. Human Communication Research, in press.

Purcell, L. & Brady, K. Assessment of interpersonal behavior in natural settings: A research technique and manual. Denver, Colo.: Children's Ashman Research Ist., 1965.

Quastler, H. A primer on information theory. In H.P. Yockey, R.L. Platzman, & H. Quastler (Eds.), Symposium on information theory in biology. New York: Pergammon Press, 1958.

Raush, H.L., Barry, W.A., Hertel, R.K., Swain, M.A. Communication, conflict and marriage. San Francisco: Josey-Bass, 1974.

Riley, M.W. Sociological research: A case approach. New York: Harcourt, Brace, and World, 1963.

Robinson, W.S. The statistical measure of agreement. American Sociological Review, 1957, 22, 17-25.

Rosen, R. Some systems theoretical problems in biology. In E. Laszlo (Ed.), Relevance of General systems theory. New York: Braziller, 1972.

Sackett, G.P. The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. Osofsky (Ed.), Handbook of infant development. New York: Wiley, 1977.

Sackett, G.P. A nonparametric lag sequential analysis for studying dependency among responses in observational scoring systems. Unpublished manuscript, University of Washington, 1974.

Sanva, V. Religion and ethnicity sociocultural factors in families of schizophrenics. Psychiatry, 1961, 24, 246-265.

Scott, W.A. Reliability of content analysis: The case of nominal scale coding. Public Opinion Quarterly, 1955, 19, 321-325.

Shannon, C.E., & Weaver, W. The mathematical theory of communication. Urbana: University of Illinois Press, 1949.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Stern, D.N. Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L.A. Rosenblum (Eds.), The effect of the infant on its caregiver. New York: Wiley, 1974.

Weick, K.E. Systematic observational methods. In G. Lindzey & E. Aronson (Eds.), Handbook of social psychology (2nd ed., vol.2). Reading, Mass.: Addison-Wesley, 1968.

Wilson, E.O. Sociobiology: The new synthesis. Cambridge: Harverd University Press, 1975.

Wood, J.T. Leading in purposive discussions: A study of adaptive behavior. Communication Monographs, 1977, 44, 152-165.

Wright, H.F. Observational child study. In P.H. Mussen (Ed.), Handbook of research methods in child development. New York: Wiley, 1960.